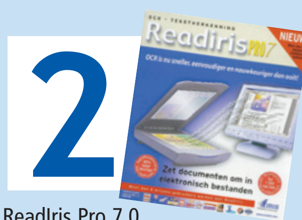


ScanSoft
OmniPage Pro 11



ReadIris Pro 7.0



ScanSoft TextBridge
Pro Millennium

OCR-pakketten

Geloof het of niet, maar je kan je pc leren lezen. Daarvoor hoef je er niet mee naar een speciaal instituut te trekken, maar je moet wel aan twee voorwaarden voldoen: je hebt een scanner nodig én een OCR-pakketje. OCR staat voor Optical Character Recognition en het betekent gewoon dat het programma een ingescande tekst niet langer als één enkel beeld waarneemt, maar dat het letters, cijfers en andere tekens los van elkaar kan herkennen …

Relatief snel na de vorige versies, pakken zowel ScanSoft als IRIS uit met vernieuwde OCR-pakketten. De derde grote speler, het Russische ABBYY, houdt het voorlopig nog bij de 'oudere' versie 5.0 van FineReader. Daarom hebben we een ander pakket van de ScanSoft-stal gehaald: TextBridge Pro Millennium.

Nu ook in pdf

Met FineReader 5.0 kon je al rechtstreeks scannen naar het Adobe Acrobat pdf-formaat. Nu is dat ook mogelijk in ReadIris Pro 7 en OmniPage Pro 11. De opties zijn zelfs wat uitgebreider. Zo maakt ReadIris een onderscheid tussen PDF Test en PDF Beeld-Tekst. Het eerste kan wel beeldmateriaal bevatten, maar het eindresultaat heeft slechts één laag. De tweede optie herkent extra items op de gescande pagina en bevat twee lagen. Daardoor is het ook mogelijk om bladwijzers aan te maken, miniaturen te creëren en achteraf door heel het pdf-document te zoeken. OmniPage gaat nog een stapje verder en biedt twee extra varianten om in pdf-formaat op te slaan, maar het zou te ver leiden om hier in detail op in te gaan. Belangrijker is dat OmniPage je in staat stelt pdf-bestanden zelf te herkennen als bron en die te converteren naar bijvoorbeeld het Word-formaat, om er eventueel wijzigingen in aan te brengen. Om het resultaat opnieuw in pdf op te slaan, heb je wel een pdf-writer nodig. Het meest voor de hand liggend is Adobe Acrobat zelf, maar er zijn ook goedkopere alternatieven.

Werken die drie nieuwe OCR-pakketten ook onder Windows XP? Uiteraard, want ze doen het wel allemaal onder Windows 2000. Toch zit er een addertje onder het gras. Als je scanner niet over de juiste TWAIN-stuurbestanden beschikt voor Windows 2000 of Windows XP, dan kan je natuurlijk niks aanvangen met OCR-programma's.



Sneller is trager

Hoe testen we OCR-programma's? Heel eenvoudig: door documenten van verschillende origine te scannen en ze via de drie OCR-pakketten te laten passeren. We gebruiken een Agfa SnapScan e40 scanner, gekoppeld aan een AMD Athlon Classic 800 MHz met 128 MB RAM onder Windows 98 SE. Het resultaat van de OCR-bewerking drukken we af en vergelijken we met het originele document. Zoals een plichtsgetrouwe leraar halen we dan de rode balpen boven en verbeteren elke fout in het document. Het gemiddelde resultaat van de verschillende documenten vormt de nauwkeurigheidsscore.

Je zal je wellicht afvragen waarom we niet opmeten hoelang een OCR-bewerking duurt. Snelheid is tenslotte een belangrijke factor bij de keuze van een OCR-programma. Wel, dat zit zo: sommige OCR-pakketten, zoals Om-

niPage Pro 11, geven je de kans om de nauwkeurigheidsgraad min of meer zelf te bepalen: ofwel sneller scannen, ofwel nauwkeuriger scannen. Hoe gek het ook mag klinken, maar als je alles in rekening brengt, is die laatste optie zelfs sneller. De tekst inscannen en herkennen verloopt inderdaad trager, maar je moet achteraf veel minder fouten manueel corrigeren. Als je voor een snellere scan kiest, is de kans op fouten beduidend hoger. Je zal de gescande tekst dan van kop tot teen moeten nalezen en dat kost extra tijd. Heb dus wat meer geduld en kies steeds voor de hoogste nauwkeurigheid. Het resultaat is niet alleen beter, maar als puntje bij paaltje komt is het ook sneller.

— Dirk Peeters —

ScanSoft OmniPage Pro 11

OmniPage Pro 11 komt vrij snel na de versie 10, maar heeft toch heel wat nieuwigheden. Over de uitgebreide pdf-ondersteuning hebben we het al gehad, maar er is ook veel aandacht besteed aan het behoud van de lay-out van het originele document. Zo worden kolommen, kop- en voetteksten, paginanummers, bijschriften en beginkapitalen behouden, om er maar enkele te noemen, al is dat niet over de hele lijn merkbaar. De intelligente nieuwe IntelliTrain-functie laat de proeflezer toe wijzingen aan te brengen en die ook onmiddellijk in de rest van het document door te voeren. Soortgelijke fouten worden op die manier automatisch hersteld. Versie 11 moet een betere verwerking van slechte originelen garanderen, zoals faxen en kopieën van kopieën. Daarvoor zorgt de Despeckle-functie die zwarte puntjes, spikkels en andere storende elementen uit het origineel filtert vóór het scanproces en de tekstherkenning beginnen. Ook tabellen zonder raster worden nu als tabel herkend. De Direct OCR-functie laat onmiddellijk tekstherkenning toe vanuit Word en Excel, zonder dat je eerst OmniPage moet starten. Het aantal herkende talen is fors toegenomen en bedraagt nu meer dan 100. Enorm handig is dat je ook langere documenten zoals boeken of verschillende losse vellen kan verwerken zonder dat je scanner over een automatische documentinvoer (ADF, Automatic Document Feeder) beschikt. Bij vele scanners is zo'n extra module namelijk niet eens aanwezig. Je kan dat nu omzeilen door een interval in te stellen tussen twee scans, om je zo de tijd te geven een andere pagina klaar te leggen.

CONCLUSIE

De nieuwe versie heeft zoveel nieuwigheden en verbeteringen dat de vorige editie (10) bij nader inzien niets meer was dan een overgangsproduct. OmniPage Pro 11 is duidelijk het pakket met de meest overzichtelijke interface en het biedt bovendien de meeste functies.

WIE-VAN-DE-3-SCORE

Nauwkeurigheid:
Interface:
Functies:
Prijsscore:
Wie-van-de-3-score: 76%

PRODUCTINFO
Product: OmniPage Pro 11
Producent: ScanSoft, 0031/36/530.25.02
Url: [www.scansoft.com]
Prijs: € 175

ReadIris Pro 7.0

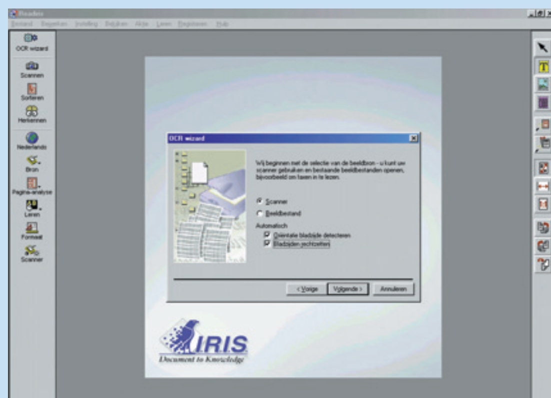
Op het eerste gezicht zien we weinig verschil met versie 6.0. Aan de interface is namelijk nauwelijks geraakt. Versie 6 was eigenlijk al een enorme verbetering ten opzichte van haar voorganger. Toch beweert IRIS dat de nauwkeurigheid met maar liefst 40% is toegenomen. Dat lijkt spectaculair, maar als je al een nauwkeurigheidsgraad claimde van meer dan 98%, dan is die 40% plots een veel minder doorslaggevend argument. Zoals FineReader en OmniPage Pro 11 kan ReadIris Pro 7 nu ook documenten in pdf-formaat bewaren. Samen met de ondersteuning van digitale camera's en de detectie van de paginaoriëntatie, is dat de belangrijkste innovatie. De ondersteuning van digitale camera's mag vreemd lijken voor een OCR-programma, maar in archiveringskringen zal die functie zeker worden gewaardeerd. Verder zijn een heleboel functies uitgebreid of gecorrigeerd. Dat is bijvoorbeeld het geval voor de herkenning van gekleurde documenten of tabellen en de eliminatie van vlekken in bijvoorbeeld faxdocumenten. Het aantal ondersteunde talen werd uitgebreid en ook de OCR Wizard is onder handen genomen. Voor wie dikwijls meerledige documenten scant, is er goed nieuws. Die mogen nu groter zijn dan

VAKTAAL

Pdf: Afkorting van portable document format. Een bestandsformaat voor opgemaakte teksten, inclusief illustraties, inhoudsopgave, verwijzingen, enz. Pdf-bestanden kan je bekijken met de Acrobatviewer, een gratis hulpprogramma van Adobe.

Twain: Stukje software dat gestandaardiseerde uitwisseling van beelden mogelijk maakt. Hierdoor kunnen bijvoorbeeld door scanners en digitale camera's gemaakte afbeeldingen worden ingelezen in elk beeldbewerkingspakket dat TWAIN ondersteunt (gelukkig doen ze dat tegenwoordig bijna allemaal).

honderd pagina's. Je moet dan wel over een scanner beschikken met automatische documentinvoer (ADF). De oplossing die OmniPage Pro 11 biedt om ook meervoudige documenten te scannen zonder ADF, is er niet. Misschien een tip voor de volgende versie?



CONCLUSIE

ReadIris was nog niet zo lang geleden het zwakkere broertje onder de OCR-pakketten. Sinds de revolutionaire versie 6.0 heeft IRIS met dit imago afgerekend. De nieuwe versie 7.0 gaat op dat elan verder en heeft enkele unieke functies, zoals de ondersteuning van digitale camera's.

WIE-VAN-DE-3-SCORE

Nauwkeurigheid:
Interface:
Functies:
Prijsscore:
Wie-van-de-3-score: 73%

PRODUCTINFO

Product: ReadIris Pro 7.0
Producent: I.R.I.S, 010/45.13.64
Url: [www.irislink.com]
Prijs: € 190 (toevoeging Aziatische talen, € 120)

ScanSoft TextBridge Pro Millennium



Moet het allemaal niet zo uitgebreid zijn en vooral, veel goedkoper, dan is TextBridge Pro Millennium misschien iets voor jou. TextBridge was het oorspronkelijke OCR-pakket van ScanSoft. Er is nooit een Nederlandse versie van gemaakt en dat is jammer, want dit pakket is specifiek voor de consument bedoeld. OmniPage Pro 11 is namelijk flink wat duurder en wordt door ScanSoft vooral afgestemd op de veeleisende, semi-professionele gebruiker. De interface is vrij eenvoudig, al vinden we er toch heel wat functies van de duurdere pakketten in terug. Maar het zijn net de nieuwigheden van OmniPage Pro 11 of ReadIris Pro 7.0, zoals de pdf-ondersteuning, die je zal moeten missen. Wie de OCR-functie van Office XP gebruikt, hoeft niet echt geld uit te geven aan TextBridge. De motor is namelijk identiek. De enige verschillen zijn de andere interface en de wat omvangrijkere functieset. Het OCR-onderdeel van Office XP is daarentegen netjes in het Nederlands en daardoor wellicht beter hanteerbaar.

CONCLUSIE

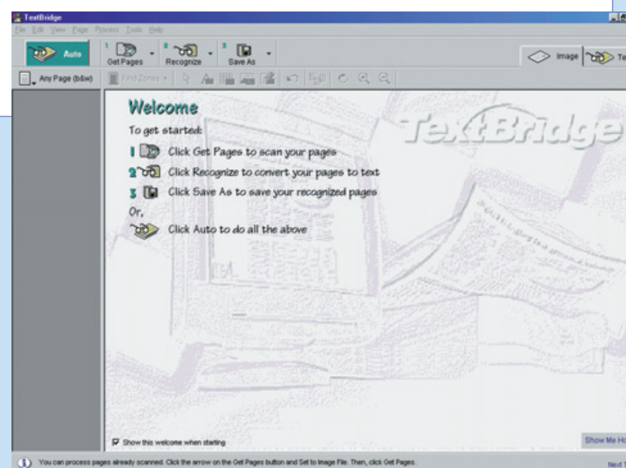
ScanSoft TextBridge Pro Millennium is een brug tussen de duurdere pakketten (ReadIris Pro 7 en OmniPage Pro 11) en de sterk afgestripte en soms verouderde OCR-programma's die je gratis bij een scanner krijgt geleverd. Je merkt het ook aan de prijs: € 75 in plaats van € 175 euro voor OmniPage. Je moet wel wat concessies doen: zo is er geen Nederlandse versie, ondersteunt TextBridge slechts 56 talen (waaronder het Nederlands) en moet je pdf-ondersteuning missen.

WIE-VAN-DE-3-SCORE

Nauwkeurigheid:
Interface:
Functies:
Prijsscore:
Wie-van-de-3-score: 81%

PRODUCTINFO

Product: TextBridge Pro Millennium
Producent: ScanSoft, 0031/36/530.25.02
Prijs: € 75
Url: www.scansoft.com



ALGEMENE CONCLUSIE

ScanSoft OmniPage Pro 11 en ReadIris 7 ontlopen elkaar nauwelijks. Op het vlak van nauwkeurigheid scoren ze even goed. Eigenlijk vinden we de interface van OmniPage net iets beter en daardoor haalt dit pakket het met een neuslengte van ReadIris. Deze beoordeling is een beetje subjectief en het kan best zijn dat jij net het omgekeerde ondervindt. Duidelijk is wel dat TextBridge het mindere broertje is, ondanks de erg aantrekkelijke prijs. Het is enkel en alleen dankzij die prijs dat het product de hoogste score haalt, want de tekstherkenning staat op een beduidend lager niveau. En ABBYY FineReader 5.0? Dat pakket kwam niet aanmerking, maar omwille van het aardig prijskaartje en de goede prestaties, blijft ook dat een aanrader.